DOCUMENT RESUME

ED 110 040 IR 002 328

AUTHOR	Bartos, Bruce
TITLE	Microform-Based Information Storage and Retrieval Systems.
INSTITUTION	Southwest Regional Laboratory for Educational Research and Development, Los Alamitos, Calif.
REPORT NO	SWRL-TM-5-71-05
PUB DATE	15 Apr 71
NOTE	36p.
EDRS PRICE	MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS	Computer Output Microfilm; *Computer Programs; Data Bases; Data Processing; Design; Electronic Data Processing; Equipment; Flow Charts; Information Dissemination; *Information Retrieval; Information Services; *Information Storage; Information Systems; Microfiche; Microfilm; *Microforms; *Systems Development
IDENTIFIERS	AIDS; *Archival Information Dissemination System

ABSTRACT

The development of microform-based information storage and retrieval systems is discussed, and the comparative advantages and costs of various approaches are considered. A proposed computer system--Archival Information Dissemination System (AIDS)--to support microform retrieval is outlined. Included are diagrams of sequential flows of information into and out of the system. Finally, the characteristics of existing computer programs for information searching are also presented. (DGC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED110049

SOUTHWEST REGIONAL LABORATORY
TECHNICAL MEMORANDUM

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

DATE: April 19, 1971

NO: TM 5-71-05

TITLE: MICROFORM-BASED INFORMATION STORAGE AND RETRIEVAL SYSTEMS

AUTHOR: Bruce Bartos

ABSTRACT

The primary concern in this paper was to reveal the information already collected for developing an information storage and retrieval system. Since many systems have to start from a hardcopy filing method, the initial emphasis was on costs of converting to microform to create a more usable data base. Next, several other data bases were discussed along with equipment available for randomly accessing information in different microforms. A generalized model for a system to help researchers find information was outlined, along with fairly detailed event schedules. The proposed system is called AIDS - An Archival Information Dissemination System. Finally, it was appropriate to cite the characteristics of several computer programs which have been written to increase the speed and accuracy of information searches. These programs can be used independently of the form (hardcopy, microfilm, or even video cassettes) of the data base.

ER008328

This document is intended for internal staff distribution and use. Permission to reprint or quote from this working document, wholly or in part, should be obtained from SWRL, 11300 La Cienega Boulevard, Inglewood, California, 90304.

MICROFORM-BASED INFORMATION STORAGE AND RETRIEVAL SYSTEMS

INTRODUCTION

This paper covers the procedural development of a microform-based information storage and retrieval system. It includes trade-off factors, costs, and benefits. The microform medium, its associated hardware, and the data bases that could be used are also covered.

LOCAL AND NATIONAL DATA BASES

Large groupings of information with at least one characteristic common across all items in any given group can be considered a data base. Most frequently these items are test scores or biographical questionnaire responses. However, the whole span of bibliographical information (books, reports lists) remain virtually untouched by automated storage, retrieval, and dissemination methodology.

If the amount of information and material increases while facilities remain constant, the available storage space usually decreases. One solution to the storage problem is to destroy that which is lowest in importance. A danger here is that what is useless today could be very necessary tomorrow. A better solution is to change the storage medium to one which provides a significant volume reduction. Microfilm provides the best of both worlds. Because of the low cost per image (approximately 1¢), large files can be converted and stored in safe, convenient locations economically. Reduction in size, however, does increase the possibility of misplacing documents. This, in turn, requires an improved filing system.

When the total data base is not too large, a file drawer system with hanging folders serves very well. But, as the base grows and information searches require several cross-reference codings, the process becomes cumbersome at best. Easing this problem can be accomplished through various automatic microfilm retrieval devices. Still, an efficient information system needs a standardized set of codes under which large groupings of data can be stored and selectively retrieved.

Since practically all publication-oriented organizations produce microfilm copies of their materials, these, when properly coded, can also become data bases. Ranging from the very structured Dissertation Abstracts of University Microfilms to the edition-theme and annual-index-by-title methods of the Phi Delta Kappan, one can find all levels of implementation in this regard. On a national, and even international, scope there are other groups which specialize in collecting, microfilming coding, and disseminating non-dissertation educational research reports. USOE's Educational Resources Information Centers (ERIC) and Phi Delta Kappa's School Research Information Service (SRIS) are outstanding examples here.

As of February, 1971, ERIC had 43,786 complete documents microfilmed and encoded in its publication, Research In Education (RIE). Similarly, almost 30,000 journal articles have been abstracted and encoded in Current Index to Journals In Education (CIJE). The RIE documents require over 65,000 microfiche, enough to fill five 4-foot file cabinets. CIJE has not been put on film yet, but at the same reduction ratio it will require over 1000 fiche. RIE and CIJE are both contained in the SWRL Library. The other five ERIC collections include: Office of Education Reports,

1956-1965 covering research reports received prior to publication of RIE; Pacesetters in Innovation is five years of ESEA Title III projects to advance creativity in education; Manpower Research is the abstracts of five years of projects supported by the Interagency Committee on Manpower Research; ERIC Catalog of Selected Documents on the Disadvantaged - needs of the disadvantaged, through 1965; and Selected Documents in Higher Education, a special grouping of 845 documents on college and university administration. These figures are cited to draw attention to the untapped potential of the ERIC system.

MICROFILM

Microfilm is fairly well standardized on raw film sizes of 16mm, 35mm and 105mm in 6 inch, 100, and 200 foot lengths. Common reduction ratios range between 15:1 and 42:1. This means from 50 images at 15:1 reduction to 224 images at 42:1 can fit on one 4 x 6 inch microfiche. Considerably beyond these ratios is the National Cash Register "Ultrafiche" system utilizing 150:1 reductions and allowing 3200 pages to be recorded on the same fiche. These are linear reductions in both the vertical and horizontal directions; the area reduction is 22,500 square units (150 x 150 = 22,500). This one fiche equals 18 inches of file drawer paper, a significant space saving.

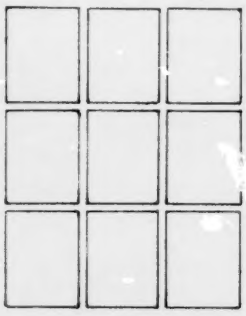
While there are endless varieties of forms of finished film, the three most popular forms are roll (cassette and open-reel), single-frame aperture card (usually 35mm), and microfiche (multiple images on one 4 x 6 inch sheet). Figures 1 and 2 give many of the combinations and permutations available.

The format used is of special importance to those interested in the microfilm aspects of ERIC both for personal use and as part of a larger information storage and retrieval system. ERIC adopted the COSATI (Committee On Scientific And Technical Information) standard of 20:1 reduction on 4 x 6 inch microfiche. This allows approximately 60 images of information per fiche, in five rows of 12 images each. USOE has even gone to the extent of recommending a particular reader which is completely compatible with its product. This reader is the DASA Corporation's PMR/50 which sells for $90. An excellent guide to all of this is the USOE publication number OE-12037-c, How To Use ERIC, for 35¢.

Along with the development of microfilm as a widely-used medium, there has also developed a special vocabulary, a verbal shorthand so to speak. Appendix A is a glossary for microforming, adapted from Modern Data, November 1970, pages 90-93.

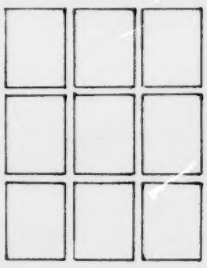
Hardcopy Conversion to Microfilm

Experience has shown that conversion of large hardcopy files to a microfilmed information storage and retrieval system is a very complex process involving interrelated manpower, software, and hardware commitments. Some of the hardware aspects were covered under the section "Comparative Features of Microform Retrieval Devices;" others will be treated in the "Archival Information Dissemination System" section. Manpower and software concerns in copy conversion are treated in this section. Computer programs (software) and currently operational search systems will be discussed under "Computer Programs and Operational



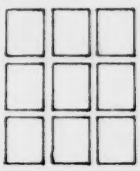
Committe on Scientific and Technical Information (COSATI)

20:1 Reduction, Avg. 60 images
Image Dimensions
10.8 mm x 14 mm



National Microfilm Association (NMA)

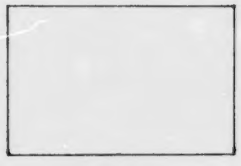
24:1 Ratio, Avg. 98 images
Image Dimensions
9 mm x 11.6 mm



New Popular

42:1 Ratio, Avg. 224 images
Image Dimensions
[illegible] mm x 6.7 mm

4" x 6" MICROFICHE
(105 mm x 148.5 mm)



APERTURE/TAB CARD

FORMAT	REDUCTION RATIO 20:1	24:1	42:1	VARIABLE
16 mm, 100' roll (max/avg)	3000/2400	3600/3000	9600/9000	NA
4"x6" microfiche (max/avg)	72/60	120/98	280/224	NA
35 mm Aperture/Tab Card	3	3	15	1

NOTE: The hard copy dimension is 8½" x 11"

Information Retrieval Systems." Costs for converting hardcopy to various microforms were received from four vendors. These are tabled in Appendix B. The names, addresses, and contact persons are found in Appendix C. Using the SWRL Archives as an example, the cost of converting 150,000 hardcopy pages to black and white silver halide 16mm roll film, at a 24:1 reduction, was requested, with all to be done in less than four weeks. These values were set because each specification has a feasible alternative which may suit one vendor better than another. For example, by setting a four-week time limit, manpower decisions can be made whether to use SWRL clerical help, or to pay the vendor price of over $5 per hour to pull staples, cut binders, straighten pages, and generally prepare the documents for microfilming. Another variable is the raw film chemical processing. Besides the specified silver halide process of black and white photography, there is the ultraviolet exposure with heat-developing Kalvar Process and the ultraviolet exposure with ammonia-developing Diazo Process. Each gives specifically characteristic results. For viewing purposes Kalvar images can stand projection lamp heat much longer than Diazo film due to the fact that Kalvar reflects light in the opaque areas rather than absorb it.

After applying a set of criteria which includes price, quality, and maximum convenience, the vendors have been rated as follows:

(1) Data Dissemination Systems Inc. (DDSI)
(2) Downey Microfilming Service (D.M.S.)/ Syntronics Computing Inc.
(3) Dymat Microforms
(4) Computer Micrographics Inc. (CMI)

A brief profile on each company is provided.

DDSI - This company offers the lowest prices for the filming and the maximum convenience for information and service.

They are representatives for several microform equipment manufacturers, including 3M Company and Morgan Information Systems. They are locally situated in Hawthorne. While no samples of their regular camera work has been inspected, several samples of COM of excellent quality were generated from our tapes and in our presence. Unlike the D.M.S./ Syntronics coalition, DDSI does not offer regular electronic data processing service bureau functions, except where directly connected to a COM contract.

D.M.S./SYNTRONICS - This represents a two-company service cooperative. D.M.S. handles all hardcopy to microform conversions and duplication, while Syntronics provides the computer services and COM customer needs. Among the joint accounts of D.M.S./Syntronics are several large continuing contracts with Max Factor Cosmetics and NCR. D.M.S.'s latest bid put them in third place price-wise. When this fact is coupled with a lack of experience or knowledge about microform retrieval equipment, they have to be rated lower than D.D.S.I.

DYMAT MICROFORMS - Dymat's prices are just slightly lower than DDSI, but they do not offer COM nor do they handle information retrieval equipment. Also, they are located in Sherman Oaks, which makes them a little less convenient for handling any of the unforeseen but inevitable problems that will arise.

CMI - This company is both a manufacturer's representative and a microforms service bureau. Their pricing technique indicates that they are very active in this area, but their cost to us is the highest of the four by several hundred dollars. In the COM area they are second lowest. They are conveniently located in the L.A. Airport area.

Once the database is in microfilm form, the most convenient and usually most economical information retrieval methods employ random access devices.

COMPARATIVE FEATURES OF MICROFORM RETRIEVAL DEVICES

Five microform retrieval devices have been investigated. HF Image Systems' CARD and Remington Rand's REMKARD are the only two devices in production at this time. However, some of the features planned for future devices might warrant postponing a decision. Appendix D presents

the most significant features of each device investigated in tabular form. The features are discussed below in light of their influence on further development of a complete information system.

HF IMAGE SYSTEMS - This device requires 105mm microfiche (or jackets). It can hold 750 microfiche which, at 24:1 reduction, gives a capacity of 73,500 images. Each fiche must have a special metal coding strip permanently attached to work in the device. This strip prevents using another viewer or reader-printer for hardcopy. Microfiche with these strips cost $4.50 each. There are over 5000 CARD devices in use world-wide. It has a computer interface option.

REMINGTON RAND - The REMKARD is a re-labelled HF Image Systems' CARD. To all appearances it is identical, with the exception of a larger viewing screen.

MICROGRAPHIC TECHNOLOGY CORPORATION - This one will also use microfiche but in a 30-fiche cartridge. The fiche can be 20X, 24X, or 42X, giving a maximum capacity of 6700 images stored in the machine. The $2000 price tag is less than half the other devices. It does not require special coding on the fiche. It does not have a computer interface. Production is slated for March, 1971. Also, MTC is marketing a microfiche camera-processor for $10,000, or one-half the cost of its closest competitor. Fully developed, ready-to-use fiche are outputted two minutes after the last image is shot.

MORGAN INFORMATION SYSTEMS INC. - The unique feature here (and its biggest stumbling block) is that is uses 105mm microfiche <u>rolls</u>. At 42X this provides a capacity of almost ½-million images, more than enough for most projected needs. Getting roll fiche made, however, is expensive and limited to only a few service bureaus. It has a $2000 16mm update option which allows for revision of specific images without throwing away the whole roll. A computer interface option is also available.

ALPHA VECTOR INC. - This company's random access reader can store 30 rolls of either 16mm roll film or 105mm roll fiche. Maximum image capacity exceeds one million images. Location coding can be EBCDIC, ASCII, or a special combination of both, making for ideal computer compatibility when the interface is developed (1972). With roll film the cost of replacing obsolete individual images, excluding labor, is

very low (less than one cent). Whole rolls cost about six dollars, comparable to other film prices. Also, rolls of colored images for the Instructional Concepts Program, Teacher Training, and Kindergarten Art Program can be interspersed among black and white "reading" rolls. The projected price for this device is $4000, slightly cheaper than three of the other devices.

A list of vendor addresses and contact personnel can be found in Appendix E.

ARCHIVAL INFORMATION DISSEMINATION SYSTEM (AIDS)

The word "archival" has been chosen so as to refer to any collection of historical documentation. Under this rubric the SWRL Archives, the ERIC files, Psychological Abstracts, or any number of other databases would be applicable. The only difference in the applicability of AIDS would be the starting point on the system development scale. For this presentation the prime assumption is that the user's organization produces hardcopy records of its work, but that it has never developed an efficient storage and retrieval system which could utilize the microfilm medium.

In order to optimize utilization of a large hardcopy file, it is advisable to alter it through four stages of development. The first stage is an information gathering and system defining task. This paper serves as a typical output from this stage. The second stage requires a reorganization of the materials to increase their use in a manual (non-mechanized) mode which will also serve as a training ground for more advanced modes. The actual microfilming and installation of automatic retrieval equipment are essential components of the third stage. Computerization, automated searches for answers, user profile matching, and remote location accessibility are projected for the fourth stage.

The detailed steps herein deal only with stages two and three, sequenced together. It is possible to stop at the end of either stage and still have an efficient system. Complete computerization is too far in the future for it to be meaningfully discussed at this time. However, there are some operational systems like GIPSY in Oklahoma, and DIALOG in San Francisco which can serve as models.

Whether completing one stage or four, cooperation between member groups is a necessity. The computer center staff, the library staff, and the documentation editing staff are the main interacters in operationalizing a system. Sometimes these groups operate autonomously within an organization as services. In other cases, the groups are a part of ongoing functional divisions. The flowcharts reflect the latter arrangement. However, the step-by-step task schedule is a merger of all events that should take place. A prose description of the entire process is also provided. While it does not have the subscript symbols which would definitively link it to the flowcharts, a quick comparison will show that they are indeed coordinated.

Event Schedule and Flowchart Details

Flowchart I shows the development from the point of forming a System Improvement Team through the point of issuing an AIDS User Guide - Manual Mode. Flowchart II starts by closing the document files in preparation for microfilming and semi-automation, and ends with a new Guide and a Technical Note on the whole development to this point. The flowcharts reflect Stages 2 and 3 segments of the complete system,

including events for both the Computer Center (CC) and the Librarian/Editor (L/E) groups. Computer Center events are identified by numeral subscripts following the event symbols, while the Librarian/Editor events have letter subscripts. These subscripts are consonant with the attached Task Schedule. Also, the following paragraphs give more detailed explanations of each CC group event; L/E is covered by inference.

Establishing a System Improvement Team will require the maximum cooperation from the L/E because they will supply four of the six members; CC will have the other two. The functions to be fulfilled will be : librarian, editor, clerk, typist, programmer-analyst, and information retrieval specialist, who should probably be the Team Supervisor.

The Team Supervisor will monitor the compilation of an extensive list of descriptors by the editor and librarian. This should be their task since they are the only ones who cross all programs from a documentation point of view. One guide can be found in the ERIC Thesaurus, which is the heart of the USOE information retrieval system.

A Team meeting will be held upon completion of the list of descriptors. The project goals, member functions and responsibilities, and review of the list will be on the agenda.

With the descriptors in hand, CC will finalize the codes and formats for the document accession numbers. It will probably be in the form: "aa-d-seq-pg." Purposely, this format is compatible with ERIC, allowing possible future coordination. "aa" stands for two alphabetic characters representing the major descriptor the particular document falls under. Filing by this descriptor will allow easy browsing. "d" is the code

number for the Division originating the document. "seq" is the chronologically-filed-within-descriptor number, up to a maximum of 999 documents. Again, this allows for focused browsing, but now by date. "pg" means number of pages, up to 99. This is included on the premise that the size of the document is an indicator of the depth of treatment of the major descriptor. The total result will be a discrete accession number which is compatible with other code formats, and which also contains unique, meaningful information.

A Management review of progress to this point can be held while the typist readies her materials for the typing of the accession numbers, titles, and descriptors on each document. The tape output from the automatic typewriter will be the input to the KWIC (Key Word In Context) index. As this typing is done, the documents will be refiled under the new retrieval scheme.

Since the typing will take several weeks to complete, the writing (coding) of a KWIC index computer program will be started. The actual coding, testing, and debugging of that program will also consume several weeks. Hopefully, these events will run in parallel.

Copies of the KWIC index will be printed and distributed to the Team with instructions for an error check. The result will be reviewed in a Team meeting before CC modifies the program and tapes. A new, final printing will almost certainly be required.

In order to keep the KWIC up-to-date in succeeding years, a KWIC update program will be written by the programmer-analyst. While this is being done, the editor will be writing the AIDS User Guide - Manual Mode, which will be of immediate utility to the organization's staff.

A Team review, and a Management review will end this stage of development.

Following Management approval of the preceding manual mode stage, the Team Supervisor will initiate the semi-automated phase with a memo to the librarian to prepare the Files for microfilming. This will require that all documents be in their proper locations, and have all necessary coding completed. Since the library staff will operate the System, they should select the vendor to do the actual microfilming, but the clerk will do the document preparation of removing staples, cutting bindings, and culling the duplicates.

During the microfilming process, the key pages (abstracts) will be saved for re-filming in a form compatible with the random access device. If this is done in the 16mm roll, duplicate sets for under $25 each could be set up at locations remote from the central machinery. Some devices have enough capacity to store the very large files (over 1 million images), thereby eliminating the need for abstracts. Rapid access (under 5 seconds for any given page), maximum file integrity (a duplicate set will be stored in a security safe off premises), and a giant step toward full automatic retrieval of information are three benefits from such devices.

A Team review will be held to assess Task progress and to pool the information for a summary document on hardware selections. The random access device, a new reader-printer, some new microfilm readers, and other needed components for the AIDS will be included.

This information will be sent to Management for review, and with their approval the equipment will be ordered.

Although attempts will be made to avoid this through earlier decisions, it is possible that additional accession coding will be necessary after the new equipment is selected. If so, the KWIC index program will have to be revised.

Likewise, the AIDS User Guide will have to be modified to include the semi-automated aspects of this system. Much of the information will be supplied by information retrieval specialists.

After the usual Team review, and an extensive Management review, a Technical Note on the entire development of the Archival Information Dissemination System will be issued.

TASK SCHEDULE

TASK - Develop an Archival Information Dissemination System (AIDS)

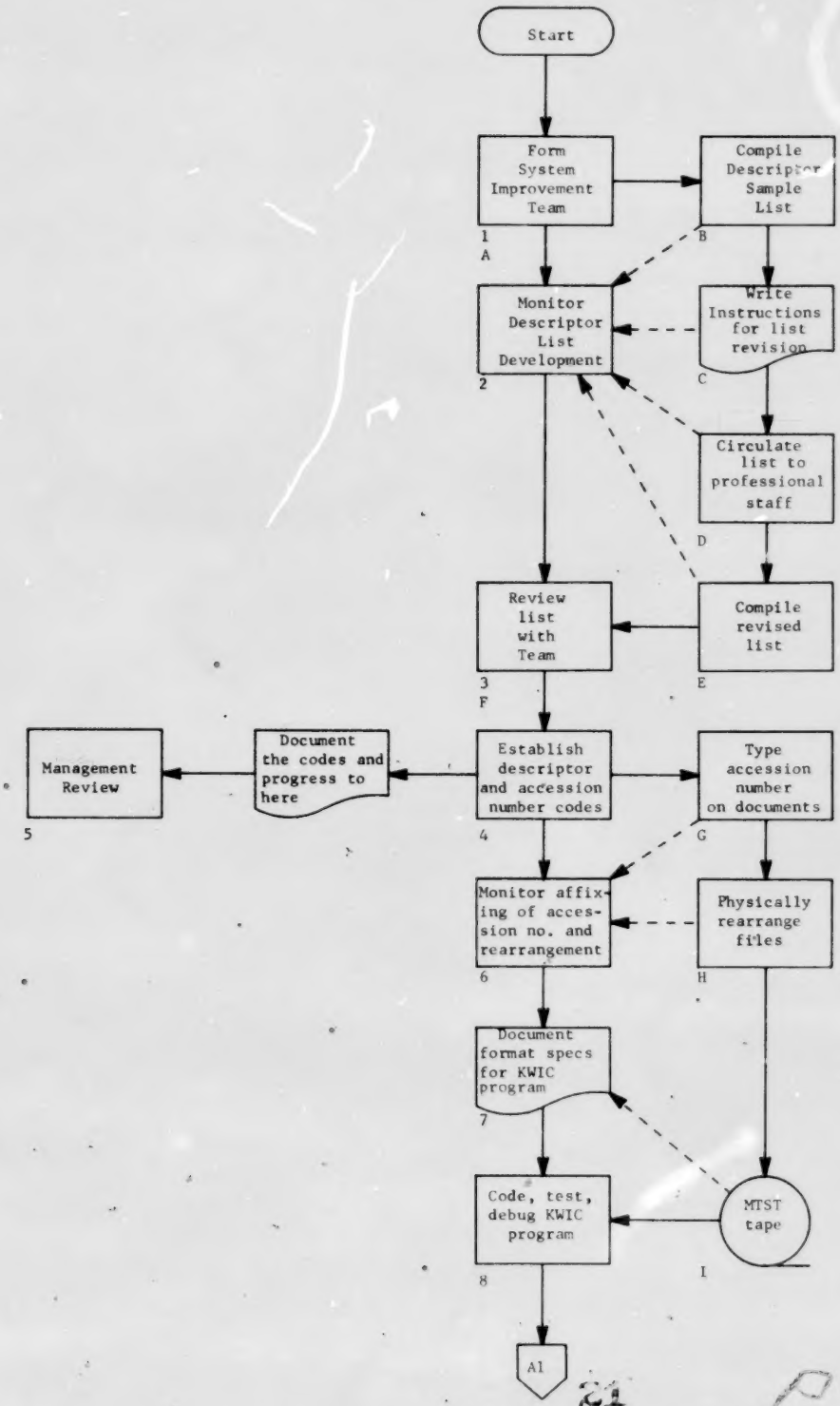
EVENT NO.	EVENT	MAN DAYS REQUIRED	STAFF
1,A.	Document job functions for members of a System Improvement Team, assign personnel, and obtain Event Schedules	3	Information Retrieval Specialist/ Management
2.	Monitor development of descriptor list	10	Info. Retr. Spec.
B.	Compile a 50 item list of possible descriptors for the documents		Editor/ Librarian
C.	Write set of instructions to users for revising and weighting the descriptors		Editor
D.	Circulate the list among selected users for revision and weighting relative to needs		Librarian
E.	Compile revised list based on user feedback		Editor/ Librarian
3,F.	Review descriptor list with Team	3	Team
4.	Establish and document codes for descriptors and accession number formats	5	Info. Retr. Spec./Programmer - Analyst
5.	Management review	3	Management
6.	Monitor the affixing of accession numbers, titles, descriptors to key page of each document, and write progress memo at half way point and end of event	20	Information Retrieval Specialist
G.	Type accession numbers, titles, and descriptors on each key page using an automatic typewriter		Typist
H.	Physically rearrange the documents in agreement with the main descriptor on each key page		Clerk

EVENT NO.	EVENT	MAN DAYS REQUIRED	STAFF
I.	Send tapes to Programmer-Analyst for use in printing AIDS KWIC Index		Typist
7.	Document format specifications for an AIDS KWIC index computer program	7	Info. Retr. Spec./Programmer Analyst
8.	Code, test, debug, and document an AIDS KWIC index computer program	15	Programmer Analyst
9.	Print mulitple copies of KWIC index and distribute to Team with error check instructions	10	Programmer Analyst/Info. Retr. Spec.
10.	Review results with Team and document revision list	3	Info. Retr. Spec.
11.	Modify and document the complete AIDS KWIC index computer program	8	Programmer Analyst
12.	Print complete KWIC index, with error check, and issue copies to users for familiarization	5	Programmer Analyst/Info. Retr. Spec.
13.	Design, code, test, and document an AIDS KWIC index update program	15	Programmer Analyst
14.	Monitor the writing of an AIDS User Guide - Manual Mode	15	Info. Retr. Spec.
J.	Write an AIDS User Guide - Manual Mode		Editor
15,K.	Review Guide with System Improvement Team, document and monitor changes, and issue final "Manual Mode" memo to management	5	Info. Retr. Spec./Team
16.	Management review of entire Manual Mode stage	3	Management
17.	Issue memo directing Librarian to close and check the document files in preparation for microfilming	3	Info. Retr. Spec.
18.	Monitor the preparation and microfilming of the files	25	Info. Retr. Spec.

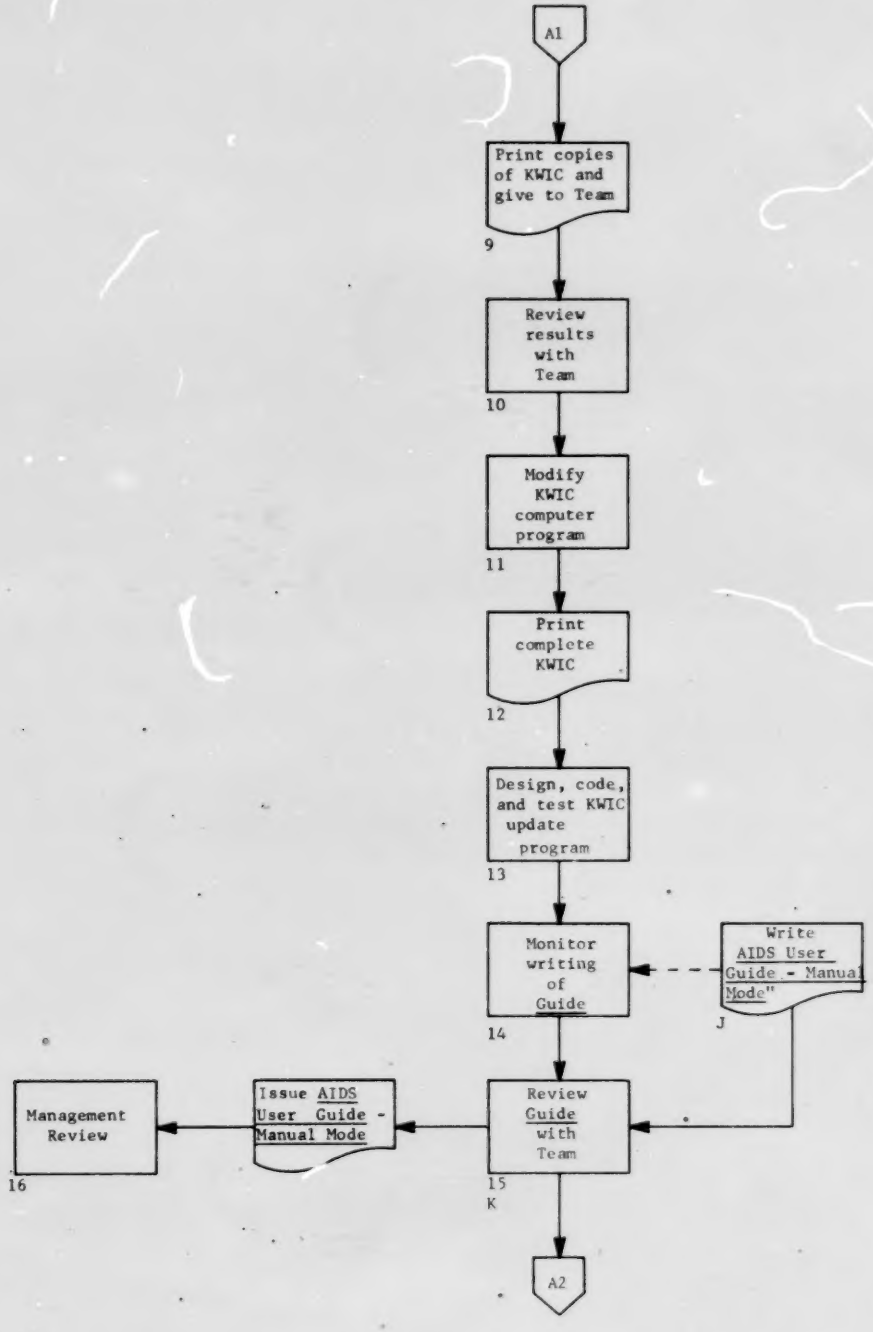
EVENT NO.	EVENT	MAN DAYS REQUIRED	STAFF
L.	Close the files and check in all missing documents		Librarian/ Clerk
M.	Select and notify the vendor doing the microfilming		Librarian
N.	Make final preparations for the microfilming, including facilities and assistants		Librarian/ Clerk
O.	Microfilm the files (whole documents)		Vendor/Clerk
P.	Microfilm the abstracts or key pages		Vendor/Clerk
19.	Select vendor for microfilm random access device and document with memo to librarian	5	Info. Retr. Spec.
Q.	Select vendor for microfilm reader-printer	5	Librarian
20,R.	Review progress with Team and issue summary document on hardware selections	5	Team
21.	Management Review	3	Management
S.	Order the microfilm reader-printer		Librarian
22.	Order the random access microfilm device	25	Info. Retr. Spec.
T.	Update the MTST tapes with random access device numbers and send tapes to Programmer Analyst		Typist
23.	Design, code, test, and document revised KWIC index computer program, to include new information from random access device	35	Programmer - Analyst
24.	Print a new AIDS KWIC index and check for errors	5	Programmer - Analyst/Info. Retr. Spec.
25,U.	Exchange information needed by users of the reader-printer and random access devices.	5	Info. Retr. Spec./Editor

EVENT NO.	EVENT	MAN DAYS REQUIRED	STAFF
V.	Write an AIDS User Guide - Semi-Automated Mode		Editor
26,W.	Review new "Guide" with Team and issue final "Semi-Automated Mode" memo to management	5	Info. Retr. Spec./Team
27.	Management review of entire Semi-Automated stage	3	Management
28.	Write Technical Report on Archival Information Dissemination System	15	Information Retr. Spec.

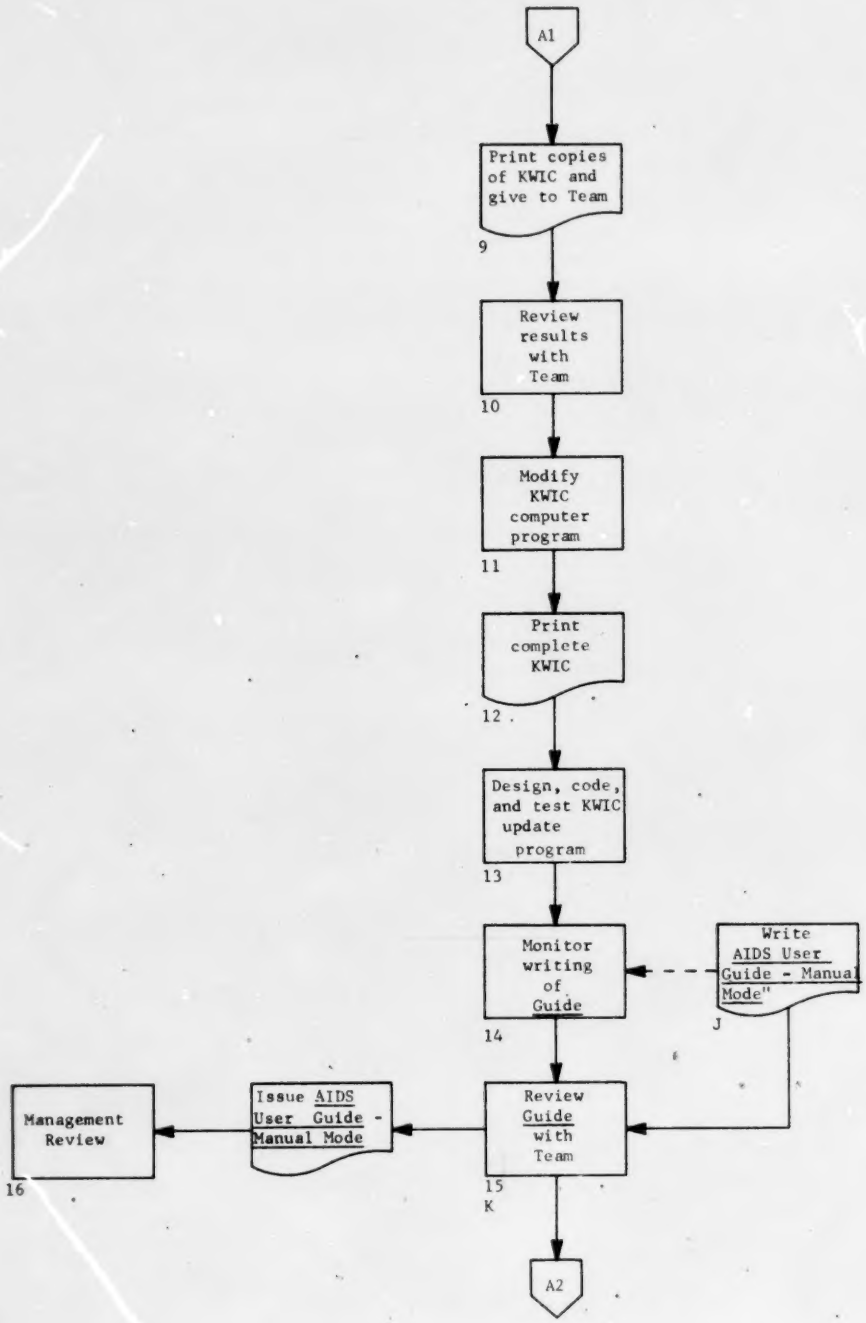
FLOWCHART I


Start
Form System Improvement Team
1
A
Compile Descriptor Sample List
B
Monitor Descriptor List Development
2
Write Instructions for list revision
C
Circulate list to professional staff
D
Review list with Team
3
F
Compile revised list
E
Management Review
5
Document the codes and progress to here
Establish descriptor and accession number codes
4
Type accession number on documents
G
Monitor affixing of accession no. and rearrangement
6
Physically rearrange files
H
Document format specs for KWIC program
7
Code, test, debug KWIC program
8
MTST tape
I
A1

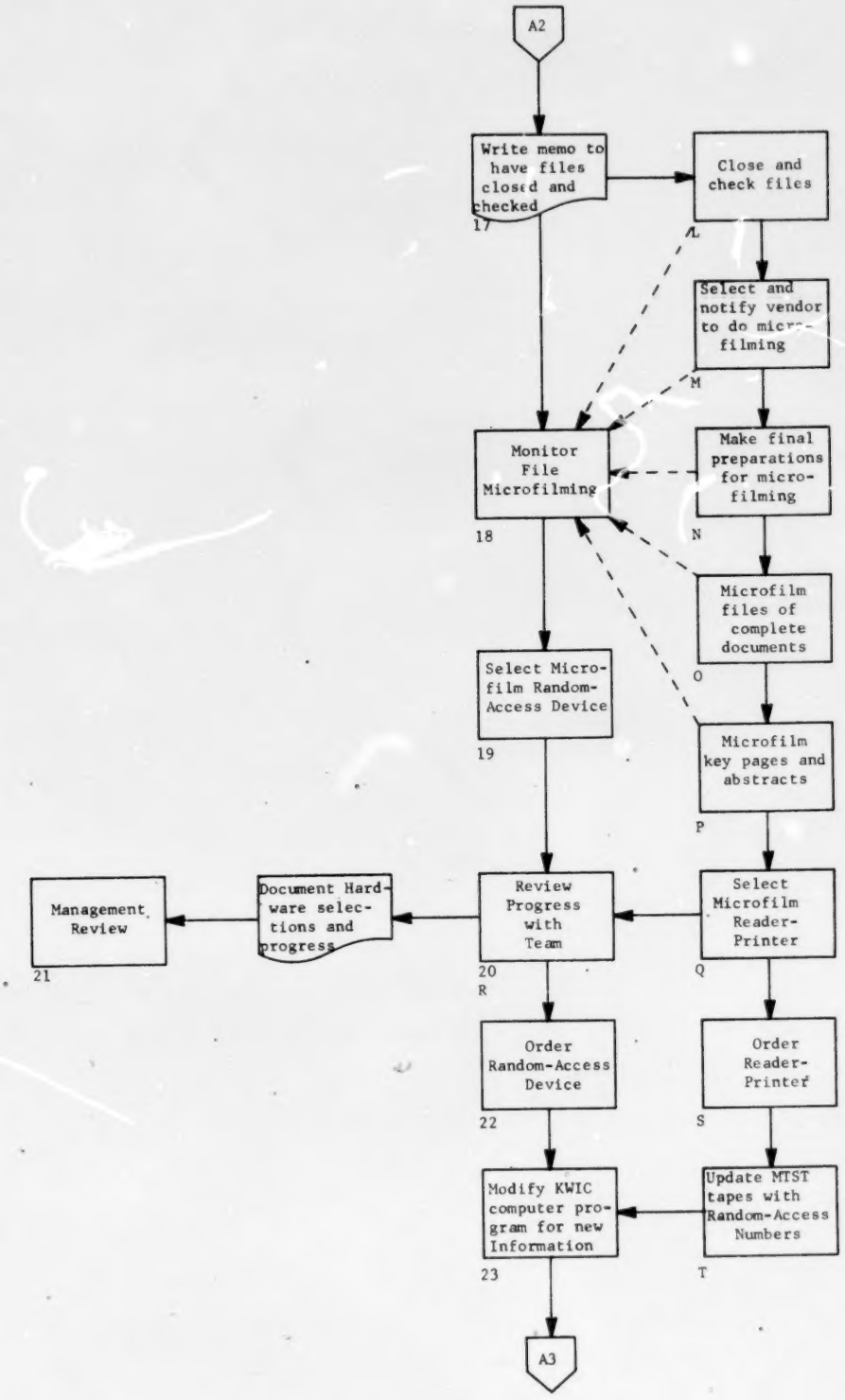
FLOWCHART I (cont.)



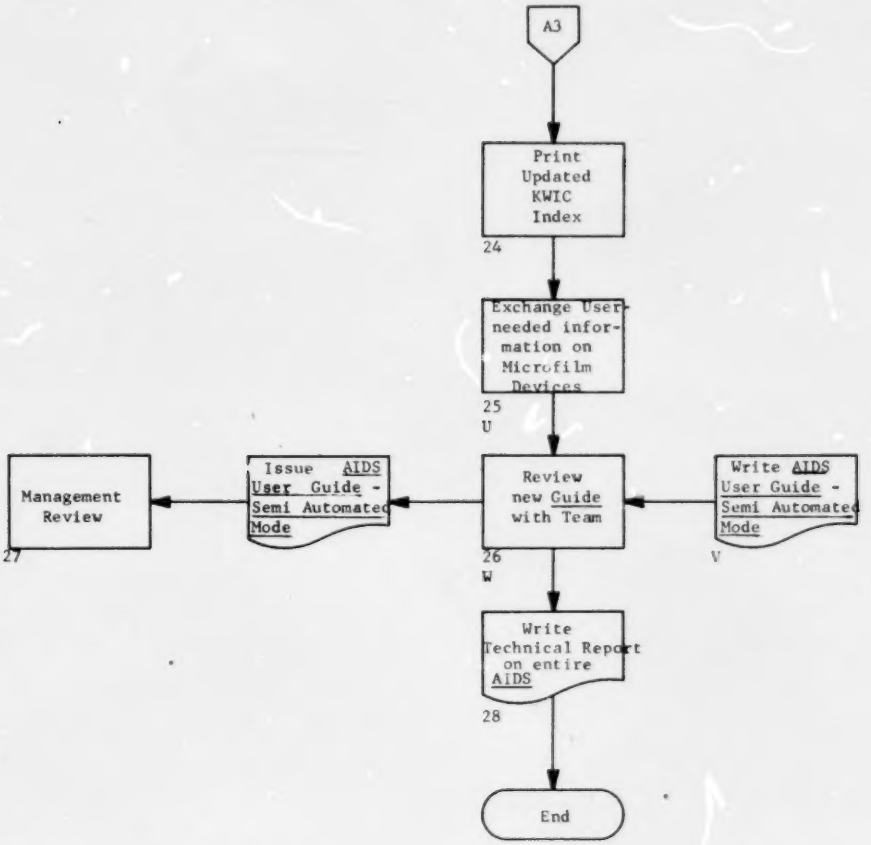
FLOWCHART I (cont.)



FLOWCHART II



FLOWCHART II (cont.)



COMPUTER PROGRAMS AND OPERATIONAL INFORMATION RETRIEVAL SYSTEMS

Because of its large size, advanced stage of development, and its Governmental support, the ERIC system has been the most popular database for which to write software packages and with which to test retrieval systems. One program which is available is called QUERY. This program is designed to search the ERIC magnetic tape file available from Leasco Systems and Research Corporation. It is written in IBM 360 Assembly Language primarily for the 360/30 system and up. However, with slight modification it is compatible with the RCA Spectra 70. QUERY is a batch processing system, hence, information requests must have the search criteria preset with logical operators of AND & OR. Batch processing is the lowest cost method, but it tends to generate more non-useful responses than an interactive system. QUERY is now installed at two ERIC Clearinghouses: Educational Media and Technology, Stanford, California; and Vocational and Technical Education, Columbus, Ohio. More specific addresses will be found in Appendix F.

The name "DIALOG" actually refers to two related, but separate, things. The first is that DIALOG is an interactive information retrieval programming language developed as a proprietary product by the Information Sciences group at Lockheed Palo Alto Research Laboratory about 1966. It was used on an IBM 360/30 with a NASA reference file of 300,000 citations. The average search time for one question was 30 minutes terminal time and 7 minutes computer time. At current time-share rates of $40 per hour for a 360/30, the answer to the search question costs $4.66. The same request handled on a 360/65 would be done in 42 seconds,

which at $300/hr. costs $3.50. On a Univac 1108, the job takes only 14 sec. but the CPU costs $800/hr., or $3.11 for the search.

The other "DIALOG" refers to the USOE utilization of the system with the ERIC files. The original installation and tryout was at the Educational Media and Technology Clearinghouse at Stanford University. After the contract ran out last Fall, however, DIALOG was moved to the Federal Office Building in San Francisco. As noted above, Stanford replaced it with QUERY, a batch processing program.

Another very useful program has been written by Aeint de Boer, a senior programmer with the UCLA Center for Information Sciences, as his Masters Thesis. It is an interactive program like DIALOG, and it is designed to access multiple databases with outputs of either lists of document numbers or the full bibliographic information including an abstract. It is written in 11 modules using PL/1 language. Although it is a rather easy thing to store and access multiple groupings of information (databases), De Boer's program is the only one which makes the claim for that capability. DIALOG was set up for NASA citations; now it has ERIC documents.

There are a variety of fully operational systems across the country. At the University of Oklahoma, the accumulated ERIC abstracts including both Research In Education and Current Index to Journals in Education are searched through the General Information Processing System (GIPSY). They will answer mail requests written in standard English. Turn-around time is about two days and abstracts cost an average of $1.70 each.

Syracuse University has put the Psychological Abstracts on-line in an interactive mode with an IBM 360/50. As of November, 1970, they were using a data base of 35,000 abstracts. The name for the system is SUPARS (Syracuse University Psychological Abstracts Retrieval Service). As of this writing, no cost per search figures are available.

The last system to be mentioned is the Science Information Exchange (SIE) in Washington, D.C. The SIE receives and stores over 100,000 reports of ongoing research in all disciplines. This is a batch processing system with, consequently, lower costs than interactive systems. For example, a request for all research associated with an individual researcher costs $2.00. Or a subject matter search on computer-assisted instruction could yield notices of 100 projects for $40.00. It currently requires two weeks to fill a search request.

CONCLUSIONS

Deciding on the best path to reach the various plateaus of a microform-based information storage and retrieval system requires taking into account the desired databases, the installation time available, the money allocated, and the current state-of-the-art of the hardware. Small, infrequently used databases warrant minimal attention; small but necessary ones cost considerably more. The dollar return on investment for accessing large files of very relevant materials should justify a substantial allocation of resources. In any case, the system designer should try to project the organization's needs for the coming five years. Manpower increases and facilities relocations are but two indicators in this direction.

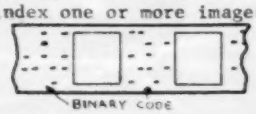
APPENDIX A

MICROFORMING GLOSSARY
(Adapted from Modern Data, Nov. 70, 90-93)

APERTURE CARD: A tab size card with a rectangular hole(s) designed to hold frame(s) of microfilm.

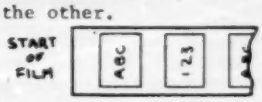
AUTOMATIC CODING: Index methods that are machine readable (i.e., digital or bit code).

BINARY DIGITAL CODE (BDC): An optical pattern of clear and opaque rectangles, machine coded for random access retrieval, used to index one or more images.



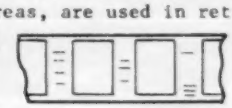
BLOWBACK: The enlargement of micro-images to readable form. Blowback ratios are expressed in diameters of magnification (a 20 X image would be magnified 20 diameters).

CINE (MOTION PICTURE (90°) ORIENTATION: Images recorded onto film in a vertical manner, one over the other.



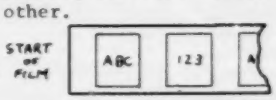
CODE AREA: Part of the recording area reserved for retrievel coding.

CODE LINE: A visual index technique where an optical pattern of clear and opaque bars, parallel with the long edge of the microfilm and located between the image areas, are used in retrieval.



COM: (1) Computer Output Microfilm - Microfilm containing data produced by computer generated signals. (2) Computer Output Microfilmer - A recorder which converts computer data into readable language onto microfilm.

COMIC STRIP (0°) ORIENTATION: Images recorded onto film in a horizontal manner, one next to the other.



CONTROL CHARACTER: A character whose occurrence in a particular context initiates, modifies, or stops a control operation.

DIAZO FILM: Copy film used to duplicate original microfilm, requiring ultraviolet light for exposure and ammonia for development.

DOCUMENT MARK (BLIP): An optical mark, usually rectangular and below the image, used for counting images or frames automatically.

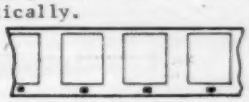


IMAGE AREA: Part of recording area reserved for image.

IMAGE CONTROL: A microfilm retrieval technique where pages are counted from the beginning of a film cartridge via the optical sensing of marks placed at the base of each image.

IMAGE MARK RETRIEVAL CODE: A document mark (blip) associated with an image used to count and locate images automatically.

JACKETS: Transparent envelopes used to simulate microfiche formatting.

KALVAR FILM: Registered name of a type of copy film used to duplicate original microfilm, requiring ultraviolet light for exposure and heat for development.

MICROFICHE: A unit record of film containing microfilm images recorded in rows or columns.

MIRACODE: Registered acronym (Microfilm Information Retrieval Access Code) for a binary digital coding technique utilizing up to 15-, 14-bar, code columns. (See also Binary Digital Code.)

PAGE (FRAME): See Code Area; Image Area; Recording Area.

RECORDING AREA: The maximum useful area containing all the recorded information, including the image and retrieval codes.

REDUCTION: An index of the size of the microfilm image as it relates to the original or blowback image. Reduction is referenced as an X number (20X denotes an image reduced 20 diameters).

VESICULAR FILM: Film sensitive to ultraviolet light and developed by heat (See also Kalvar Film; Xidex HD Film).

APPENDIX B

HARDCOPY FILE SPECIFICATIONS

150,000 pages	16mm roll film	3000 pages/roll
160 hours preparation	50 rolls of film	24:1 reduction ratio

HARDCOPY FILE CONVERSION COSTS

DDSI		
filming	50 rolls x $24/roll =	$1200.00
duplicates	50 rolls x $ 6/roll =	300.00
prep	160 hrs x $5.50/hr =	880.00
		2380.00
D.M.S./SYNTRONICS		
filming	50 rolls x $25/roll =	$1250.00
duplicates	50 rolls x $ 4/roll =	200.00
prep	160 hrs x $ 6/hr =	960.00
		2410.00
DYMAT		
filming	50 rolls x $27/roll =	$1350.00
duplicates	50 rolls x $4.50/roll =	225.00
prep	160 hrs x $ 5/hr =	800.00
		2375.00
CMI		
filming	50 rolls x $29.30/roll =	$1465.00
duplicates	50 rolls x $ 6/roll =	300.00
prep	160 hrs x $8.50/hr =	1360.00
		3125.00

APPENDIX B

HARDCOPY FILE SPECIFICATIONS

150,000 pages	16mm roll film	3000 pages/roll
160 hours preparation	50 rolls of film	24:1 reduction ratio

HARDCOPY FILE CONVERSION COSTS

DDSI			
filming	50 rolls x $24/roll =		$1200.00
duplicates	50 rolls x $ 6/roll =		300.00
prep	160 hrs x $5.50/hr =		880.00
			2380.00
D.M.S./SYNTRONICS			
filming	50 rolls x $25/roll =		$1250.00
duplicates	50 rolls x $ 4/roll =		200.00
prep	160 hrs x $ 6/hr =		960.00
			2410.00
DYMAT			
filming	50 rolls x $27/roll =		$1350.00
duplicates	50 rolls x $4.50/roll =		225.00
prep	160 hrs x $ 5/hr =		800.00
			2375.00
CMI			
filming	50 rolls x $29.30/roll =		$1465.00
duplicates	50 rolls x $ 6/roll =		300.00
prep	160 hrs x $8.50/hr =		1360.00
			3125.00

APPENDIX B (cont.)

DETAILED MICROFORM COSTS

	RED. RATIO	16MM ROLL	DUPE ROLL	PREP PER HOUR	105MM FICHE	COM FICHE	ROLL FICHE	DUPE FICHE
DDSI	24X	24.00	6.00	5.50	2.80	1.96	4.70	.24
	42X				2.80	1.96	5.00	.24
D.M.S./ SYNTRONICS	24X	25.00	4.00	6.00	Jackets w/ 72	2.80		.25
	32X	25.00	4.00	6.00	images per fiche			.25
	42X	25.00	4.00	6.00		4.80		.25
DYMAT	24X	27.00	4.50	5.00	3.86		4.00	.25
	42X			5.00			6.80	.25
CMI	24X	29.30	6.00	8.50	Jackets w/72 im @1.65/fc	2.50		.20
	42X			8.50				.20

APPENDIX C

VENDORS OF HARDCOPY TO MICROFORM CONVERSION

Computer Micrographics Inc. (CMI)
5345 West 102 Street
Los Angeles, California
(213) 776-6820
Ron Calkins

Data Dissemination Systems Inc. (DDSI)
4736 El Segundo Boulevard
Hawthorne, California 90250
(213) 644-4425
Fred Clark

Dymat Microforms
14241 Ventura Boulevard
Sherman Oaks, California 91403
(213) 783-7526
Bernie Meyer

Syntronics Computing Inc.
15130 Ventura Boulevard
Sherman Oaks, California 91403
(213) 783-6122
Ted Ross

Downey Microfilming Service (DMS)
12130 Woodruff Ave.
Downey, California 90241
(213) 923-9391
Juan Trujillo

APPENDIX D

MICROFORM RETRIEVAL DEVICES SPECIFICATIONS

	MODEL	MAGNIFI-CATION	UNIT CAPACITY	MICRO-FORM	SCREEN SIZE H x W	MACHINE H x W x D	WEIGHT POUNDS	ACCESS TIME SEC.	TOTAL IMAGE CAPACITY	COST DOLLARS	RETRIEVAL CODING	FORM HOLDER
1. HF IMAGE SYSTEMS, INCORPORATED	CARD	20x 24x	750 Fiche	Fiche	11 x 14	19/18/25	80	4	45,000 73,500	5000 basic 1500 comp. interface	Metal strip	cartridge
2. REMINGTON RAND, INC.	REMKARD	20x 24x	750 Fiche	Fiche	12 x 18	19/18/25	93	4	45,000 73,500	5000 basic 1500 comp. interface	Metal strip	cartridge
3. MICROGRAPHIC TECHNOLOGY CORPORATION		20x 24x 42x	30 Fiche	Fiche	14 x 14	20/18/24	75	3	1,800 2,940 6,720	2000	None	single and cartridge
4. MORGAN INFORMATION SYSTEMS, INCORPORATED	DUO 200	20x 24x 42x	500'	105mm roll film	11 x 14	48/32/30	185	2.5 1.0	60,000 98,000 224,000	5000 basic 2000 update 1000 comp. interface	BCD	single roll
5. ALPHA-VECTOR, INCORPORATED	AV 100	20x 24x 42x	500'	16mm 105mm roll film	15 x 12	21/18/24	80	5	400,000 500,000 1,145,000	4000	microcode Avector, EBCDIC, ASCII	Up to 30 cassettes

100', 16mm Roll Film

20x = 2400 Images
24x = 3000 "
42x = 9000 "

One 4" x 6" Fiche

20x = 60 Images
24 = 98 "
42 = 224 "

APPENDIX E

SUPPLIERS OF MICROFORM RETRIEVAL SYSTEMS

Alpha-Vector, Inc.
501 Fifth Avenue
New York, New York 10017
(212) 869-1735
Mr. Lindner

HF Image Systems, Inc.
11244 Playa Court
Culver City, California 90230
(213) 390-3378
Leon Bloom

Micrographic Technology Corporation
1732 Kaiser Avenue
Santa Ana, California 92705
(714) 540-9688
Preston Weber

Morgan Information Systems Inc.
3197 Park Boulevard
Palo Alto, California 94306
(415) 327-3991
Lew Rinehart

Data Dissemination Systems Inc. (Reps to Morgan Info. Sys.)
4736 El Segundo Boulevard
Hawthorne, California 90250
(213) 644-4425
Fred Clark

Remington Rand
1730 West Olympic Boulevard
Los Angeles, California 90015
(213) 386-1400
Art Wahlberg

APPENDIX F

COMPUTER PROGRAMS AND OPERATING INFORMATION RETRIEVAL SYSTEMS

A. DIALOG

Lockheed Palo Alto Research Lab.
Information Sciences Group
Palo Alto, Calif.
Dr. Roger Summit

ERIC Clearinghouse on Educational Media and Technology
Stanford University
Stanford, Calif. 94305
Michele Timbie

B. QUERY

ERIC
U.S. Office of Education, Room 3008
400 Maryland Avenue, S.W.
Washington, D.C. 20202
James Eller

ERIC Clearinghouse on Educational Media and Technology
Stanford University
Stanford, Calif. 94305
Michele Timbie

ERIC Information Analysis Center
for Science and Mathematics Education
1460 West Lane Avenue
Columbus, Ohio 43210
Dr. Robert W. Howe, Director

C. GIPSY

Information Systems and Evaluation Center
College of Education
820 Van Fleet Oval, Room 112
Norman, Oklahoma 73069

D. ERIC MAGTAPES

ERIC Facility
Leasco Systems and Research Corp.
4833 Rugby Avenue
Bethesda, Md. 20014

Science Information Exchange
1730 M Street, N.W.
Washington, D.C. 20036

F. School Research Information Service

Phi Delta Kappa
8th and Union Sts.
Bloomington, Ind. 47401
Dr. Neville Robertson